The Shifted QR Algorithm for Hermitian Matrices

T. J. DEKKER*
Mathematical Center
Amsterdam, The Netherlands

AND

J. F. TRAUB† ‡
Bell Telephone Laboratories, Incorporated
Murray Hill, New Jersey

Communicated by A. S. Householder

1. INTRODUCTION

Let $A = A_1$ be a given nth-order square matrix, and define a sequence of matrices by

$$A_i - \sigma_i = Q_i R_i, \tag{1.1}$$

$$A_{i+1} - \sigma_i = R_i Q_i, \tag{1.2}$$

where Q_i is unitary and R_i is upper triangular. (We suppress the identity matrix when it multiplies a scalar.) Then $A_{i+1} = Q_i^* A_i Q_i$, so the A_i are unitarily similar. If the A_i approach a form in which all elements in the last row except the lower-right element are zero, we say the A_i are *convergent*. If the A_i are convergent, the lower-right element converges to an eigenvalue of A.

The decomposition is unique if σ_i is not an eigenvalue of A and if the diagonal elements of R_i are chosen positive. The scalar σ_i, which is

* This paper is based on work done while T. J. Dekker was visiting Bell Telephone Laboratories.

† Some of the material in this paper was presented by J. F. Traub in a talk, by invitation, at the Gatlinburg Symposium on Numerical Algebra, April 1969.

‡ Present Address: University of Washington, Seattle, Washington.

Linear Algebra and Its Applications 4(1971), 137–154

used to accelerate convergence, is called the "shift parameter." The shift is determined by solving an eigenvalue problem of order k, $k < n$. In this paper we define two classes of shift strategies. First- and second-order *Class One* shifts are used in contemporary practice. Wilkinson [9] has shown that, for real symmetric tridiagonal matrices, QR is convergent if second-order Class One shifts are used. This result is false if A is not tridiagonal. We introduce *Class Two* shifts and show that, for any Hermitian matrix, QR is convergent if second-order Class Two shifts are used.

We study how certain functions transform under the QR transformation. We identify Class One and Class Two as nonstationary Newton-Raphson iteration.

In this paper we assume that A is Hermitian. We analyze non-Hermitian matrices in a later paper.

2. NOTATION AND SUMMARY

Let $E_k = E$ denote the last k columns of the identity matrix for $k \leq n$. Let the shift $\sigma_i^{(k)}$ satisfy

$$\det[E_k^*(A_i - \lambda)E_k] = 0. \tag{2.1}$$

(For uniformity we use the symbol * which denotes conjugate transpose even for real matrices.) For $k > 1$, choose the solution of (2.1) which is closest to the solution of (2.1) for $k - 1$.

We define this as a *Class One shift of order* k. The shift strategies used in contemporary practice are first- and second-order Class One shifts (Wilkinson [8]).

We shall show that QR with a Class One shift may be viewed as the following nonstationary iteration. A certain generalization of Newton-Raphson iteration is applied to a sequence of rational functions with matrix coefficients. If we define

$$v_i^{(k)}(\lambda) = v_i(\lambda) = E_k^*(A_i - \lambda)^{-1}E_k, \tag{2.2}$$

then we shall show that a Class One shift $\sigma_{i+1}^{(k)} = \sigma_{i+1}$ satisfies

$$\det[\sigma_i - \sigma_{i+1} + v_i(\sigma_i)[v_i'(\sigma_i)]^{-1}] = 0, \tag{2.3}$$

where v_i transforms according to the rule

$$v_{i+1}(\lambda) = r_i v_i(\lambda, \sigma_i, \sigma_i)[v_i'(\sigma_i)]^{-1} r_i^{-1}, \qquad r_i^{(k)} = r_i = E^* R_i E, \tag{2.4}$$

where $v_i(\lambda, \sigma_i, \sigma_i)$ denotes a second-order divided difference.

We stress that the sequence $v_i(\lambda)$ is not formed explicitly. Jenkins and Traub [4] generate an explicit nonstationary iteration to obtain a globally convergent algorithm for polynomials. Their method applies an iteration to a sequence of rational functions converging to a linear polynomial. Equation (4.7) shows how this situation may occur for QR. The study of the relation between matrix transforms and functional iteration was initiated by Bauer [1]. He considered only $k = 1$ and used techniques of analysis quite different from those we shall use.

Although the idea of shifting originated with Francis [2], the first analysis of convergence of shifted QR was published only recently by Wilkinson [9] and Parlett and Kahan [7].

If the A_i approach a form in which all elements in the last row except the lower-right element are zero, we say QR is *convergent*.

This definition of convergence differs from the definition (Parlett [6]) used for unshifted QR. We state a theorem of Wilkinson [9] in our terminology.

Let A be real symmetric tridiagonal. Then, if QR with second-order Class One shifts is used, the A_i are convergent.

The following example shows that this result is false if A is symmetric but not tridiagonal. Let

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Then, for first- or second-order Class One shifts, $A_i = A$, $\sigma_i = 0$ for all i. Parlett and Kahan [7] showed how first-order Class One shifts may fail to converge for nontridiagonal Hermitian matrices.

We introduce a new class of shifts. Let $e = E_1$ and let

$$B_i^{(k)} = B_i = (e, A_i e, \ldots, A_i^{k-1} e), \qquad k \leqslant n \tag{2.5}$$

Thus B_i is an (n, k) matrix whose columns form a "Krylov sequence" [3, p. 18]. Assume that $B_i^* B_i$ is nonsingular. Let the shift σ_i satisfy

$$\det[B_i^*(A_i - \lambda)B_i] = 0. \tag{2.6}$$

For $k > 1$, choose the solution of (2.6) which is closest to the solution of (2.6) for $k - 1$. We define this as a *Class Two shift of order k*.

Note that, if $k = 1$, $B_i{}^*B_i$ is nonsingular. If $k = 2$, $B_i{}^*B_i$ is singular only if the last column of A_i is proportional to e. This means that A_i has converged. Hence the requirement that $B_i{}^*B_i$ be nonsingular is an additional condition only for $k > 2$.

We shall show that QR with a Class Two shift may be viewed as a generalization of Newton-Raphson iteration applied to a certain sequence of rational functions. Furthermore, Class Two shifts are zeros of "Lanczos polynomials." Finally we shall prove that QR with second-order Class Two shifts is convergent for Hermitian matrices.

It is not difficult to show that, for a tridiagonal symmetric matrix, Class One shifts and Class Two shifts of the same order are identical. In numerical practice, Hermitian matrices are generally reduced to real symmetric tridiagonal form before QR is applied. Hence our convergence result is primarily of theoretical interest. It may be viewed as a coordinate-free theorem on convergence. It does, however, have practical applicability for band matrices.

3. GENERAL TRANSFORMATION RULES

We study how certain functions transform under the QR transformation. The functions include the resolvent, certain sections of the resolvent, and the characteristic polynomials of the principal minors. *The results of the section follow from the QR transformation and are independent of the shift strategy. We assume throughout that σ_i is not an eigenvalue of A.*

The resolvent of A_i is defined by

$$V_i(\lambda) = (A_i - \lambda)^{-1}. \tag{3.1}$$

Let $f(\tau_0, \tau_1, \ldots, \tau_p)$ denote the pth divided difference of f with respect to $\tau_0, \tau_1, \ldots, \tau_p$. The following basic relation is easily verified.

LEMMA 3.1. *Let $\tau_0, \tau_1, \ldots, \tau_p$ be $p + 1$ arbitrary numbers. Then*

$$V_i(\tau_0)V_i(\tau_1) \cdots V_i(\tau_p) = V_i(\tau_0, \tau_1, \ldots, \tau_p). \tag{3.2}$$

The transformation rule for $V_i(\lambda)$ is given by

THEOREM 3.1.

$$V_{i+1}(\lambda) = R_i V_i(\lambda, \sigma_i, \sigma_i)[V_i'(\sigma_i)]^{-1}R_i^{-1}. \tag{3.3}$$

Proof.

$$V_{i+1}(\lambda) = (A_{i+1} - \lambda)^{-1} = R_i V_i(\lambda) R_i^{-1}.$$

Hence

$$V_{i+1}(\lambda) = R_i V_i(\lambda) V_i(\sigma_i) Q_i.$$

From

$$V_i(\sigma_i) = V_i^*(\sigma_i) = Q_i^{-*} R_i^{-*},$$

we have

$$Q_i = V_i(\sigma_i) R_i^*.$$

Hence

$$V_{i+1}(\lambda) = R_i V_i(\lambda) V_i(\sigma_i) V_i(\sigma_i) R_i^* = R_i V_i(\lambda, \sigma_i, \sigma_i) R_i^*. \tag{3.4}$$

From

$$(A_i - \sigma_i)^{-1} = R_i^{-1} Q_i^{-1}, \qquad (A_i - \sigma_i)^{-*} = Q_i^{-*} R_i^{-*},$$

we find

$$V_i'(\sigma_i) = (R_i^* R_i)^{-1}. \tag{3.5}$$

The theorem follows from (3.4) and (3.5).

Let $1 \leqslant k \leqslant n$ and let $E_k = E$ denote the last k columns of the nth-order identity matrix. Let

$$v_i^{(k)}(\lambda) = v_i(\lambda) = E^* V_i(\lambda) E. \tag{3.6}$$

Thus $v_i(\lambda)$ is the lower-right submatrix (k, k) of the resolvent. This matrix function will play a basic role in our analysis.

We state a useful and easily verifiable result. It is implicitly stated and used in Parlett and Kahan [7].

LEMMA 3.2. *Let R be any upper triangular matrix, and let $r = E^*RE$. Then*

$$R^*E = Er^*, \tag{3.7}$$

$$E^*R = rE^*. \tag{3.8}$$

Equation (3.8) is, of course, equivalent to (3.7).
The transformation rule for $v_i(\lambda)$ is given by

THEOREM 3.2.

$$v_{i+1}(\lambda) = r_i v_i(\lambda, \sigma_i, \sigma_i)[v_i'(\sigma_i)]^{-1} r_i^{-1}, \qquad r_i = E^* R_i E. \tag{3.9}$$

Proof. From (3.4),

$$v_{i+1}(\lambda) = E^* R_i V_i(\lambda, \sigma_i, \sigma_i) R_i^* E. \tag{3.10}$$

Using Lemma 3.2,

$$v_{i+1}(\lambda) = r_i E^* V_i(\lambda, \sigma_i, \sigma_i) E r_i^* = r_i v_i(\lambda, \sigma_i, \sigma_i) r_i^*. \tag{3.11}$$

From (3.5),

$$v_i'(\sigma_i) = (r_i^* r_i)^{-1} \tag{3.12}$$

and the result follows.

Let

$$\Phi(\lambda) = \det(A_i - \lambda) = \det(A - \lambda) \tag{3.13}$$

denote the characteristic polynomial of A. Let $\varphi_i^{(k)}(\lambda) = \varphi_i(\lambda)$ denote the characteristic polynomial of the upper-left $(n - k)$-order principal submatrix of A_i. It is easy to show that

$$\varphi_i(\lambda) = \Phi(\lambda) \det v_i(\lambda). \tag{3.14}$$

From Theorem 3.2 we have that

$$\det v_{i+1}(\lambda) = \det[v_i(\lambda, \sigma_i, \sigma_i)[v_i'(\sigma_i)]^{-1}]. \tag{3.15}$$

Equations (3.14) and (3.15) establish the transformation rule for $\varphi_i(\lambda)$. For the case $k = 1$, this may be written out explicitly as

$$\varphi_{i+1}(\lambda) = \frac{1}{v_i'(\sigma_i)(\lambda - \sigma_i)^2} \{\varphi_i(\lambda) - \Phi(\lambda)[v_i(\sigma_i) + (\lambda - \sigma_i)v_i'(\sigma_i)]\}. \tag{3.16}$$

Hence, if ρ_j is an eigenvalue of A,

$$\varphi_{i+1}(\rho_j) = \frac{\varphi_i(\rho_j)}{v_i'(\sigma_i)(\rho_j - \sigma_i)^2}. \tag{3.17}$$

Assume that the eigenvalues of A are distinct. Let $v_i(\lambda) = [\varphi_i(\lambda)]/[\Phi(\lambda)]$ have the partial fraction expansion

$$v_i(\lambda) = \sum_{j=1}^{n} \frac{h_j^{(i)}}{\lambda - \rho_j}.$$

Then, from (3.15) or (3.17),

$$v_{i+1}(\lambda) = \sum_{j=1}^{n} \frac{h_j^{(i)}}{v_i'(\sigma_i)(\rho_j - \sigma_i)^2(\lambda - \rho_j)}. \tag{3.18}$$

Let

$$a_i^{(k)} = a_i = E^* A_i E. \tag{3.19}$$

The transformation rule for a_i is given by

Theorem 3.3.

$$a_{i+1} = r_i[\sigma_i + v_i(\sigma_i)[v_i'(\sigma_i)]^{-1}]r_i^{-1}. \tag{3.20}$$

Proof. From

$$A_{i+1} - \sigma_i = R_i V_i(\sigma_i) R_i^*$$

and from Lemma 3.2, it follows that

$$a_{i+1} - \sigma_i = r_i v_i(\sigma_i) r_i^*. \tag{3.21}$$

The result follows from (3.12) and (3.21).

4. CLASS ONE SHIFTS

We remind the reader that a Class One shift σ_i satisfies

$$\det[E^*(A_i - \lambda)E] = 0. \tag{4.1}$$

We shall prove that Class One shifts may always be written in terms of a nonstationary iteration involving only the function $v_i(\lambda)$. First we prove a result which relates the eigenvalues of A with the poles of $v_i(\lambda)$.

Theorem 4.1. *All poles of $v_i(\lambda)$ are eigenvalues of A. Furthermore $v_i(\lambda)$ has at least one pole.*

Proof. From

$$(A_i - \lambda)\,\mathrm{adj}(A_i - \lambda) = \Phi(\lambda)I,$$

where $\Phi(\lambda)$ denotes the characteristic polynomial of A, it follows that

$$v_i(\lambda) = \frac{E^*\,\mathrm{adj}(A_i - \lambda)E}{\Phi(\lambda)}. \tag{4.2}$$

The numerator is a polynomial of degree exactly $n - 1$ with matrix coefficients of order k. The theorem follows immediately.

We prove the major result of this section.

THEOREM 4.2. *A Class One shift* σ_{i+1} *satisfies*

$$\det[\sigma_i - \sigma_{i+1} + v_i(\sigma_i)[v_i'(\sigma_i)]^{-1}] = 0. \tag{4.3}$$

Proof. The defining equation for a Class One shift may be written as $\det[a_{i+1} - \lambda] = 0$.

From Theorem 3.3, we know that

$$a_{i+1} - \lambda = r_i[\sigma_i - \lambda + v_i(\sigma_i)[v_i'(\sigma_i)]^{-1}]r_i^{-1},$$

and the result follows.

From the theorem we find immediately

COROLLARY 4.1. *Let*

$$z_i(\lambda) = v_i^{-1}(\lambda). \tag{4.4}$$

A Class One shift satisfies

$$\det[\sigma_i - \sigma_{i+1} - [z_i'(\sigma_i)]^{-1}z_i(\sigma_i)] = 0. \tag{4.5}$$

If $k = 1$, this is just

$$\sigma_{i+1} = \sigma_i - \frac{z_i(\sigma_i)}{z_i'(\sigma_i)}, \tag{4.6}$$

which is Newton-Raphson iteration applied to the rational function $z_i(\lambda)$. From (3.14) and (4.4) we have that $z_i(\lambda)$ is a rational function whose numerator is $\Phi(\lambda)$, the characteristic polynomial of A, and whose denominator is $\varphi_i(\lambda)$, the characteristic polynomial of the upper-left $(n - 1)$-order principal submatrix of A_i.

We can obtain additional insight into the QR algorithm by the following considerations. Assume that QR is convergent and that the lower-right element converges to ρ. Then

$$z_i(\lambda) \to \rho - \lambda. \tag{4.7}$$

Hence Newton-Raphson iteration is being applied to a rational function which is converging to a linear polynomial.

In the case of arbitrary k it follows from Theorem 4.1 that all zeros of $z_i(\lambda)$ are eigenvalues of A and that $z_i(\lambda)$ has at least one zero. Equation (4.5) gives the shift as a *generalized Newton-Raphson iteration*. (Equation 4.3 gives a generalized Newton-Raphson iteration appropriate to poles.)

The matrix $z'(\lambda)$ is bounded away from zero. Indeed, we have

THEOREM 4.3. *At all points where $\varphi_i(\lambda) \neq 0$, $z_i'(\lambda) + I$ is negative definite.*

Proof. Partition A_i as

$$A_i = \begin{bmatrix} \mathscr{A}_i & b_i \\ b_i^* & a_i \end{bmatrix},$$

where a_i is a (k, k) matrix and $\mathscr{A}_i$ is a $(n-k, n-k)$ matrix. Then it is easy to show that

$$z_i(\lambda) = a_i - \lambda - b_i^*(\mathscr{A}_i - \lambda)^{-1}b_i,$$

where $(\mathscr{A}_i - \lambda)^{-1}$ is nonsingular by hypothesis. Hence

$$z_i'(\lambda) + I = -[(\mathscr{A}_i - \lambda)^{-1}b_i]^*(\mathscr{A}_i - \lambda)^{-1}b_i,$$

and the result follows.

From Theorems 3.2 and 4.2 we can summarize as follows. The QR algorithm with Class One shifts is the iteration

$$\begin{aligned} &\det[\sigma_i - \sigma_{i+1} + v_i(\sigma_i)[v_i'(\sigma_i)]^{-1}] = 0, \\ &v_{i+1}(\lambda) = r_i v_i(\lambda, \sigma_i, \sigma_i)[v_i'(\sigma_i)]^{-1} r_i^{-1}. \end{aligned} \tag{4.8}$$

For $k = 1$,

$$\sigma_{i+1} = \sigma_i + \frac{v_i(\sigma_i)}{v_i'(\sigma_i)}, \qquad v_{i+1}(\lambda) = \frac{v_i(\lambda, \sigma_i, \sigma_i)}{v_i'(\sigma_i)}. \tag{4.9}$$

5. CLASS TWO SHIFTS

We remind the reader that a Class Two shift σ_i satisfies

$$\det[B_i^*(A_i - \lambda)B_i] = 0, \tag{5.1}$$

where $B_i = (e, A_ie, \ldots, A_i^{k-1}e)$ and where $B_i^*B_i$ is assumed nonsingular.

We prove a lemma which is analogous to Lemma 3.2.

Lemma 5.1. *Let $\{R_i\}$ be the sequence of upper triangular matrices generated by the QR transform. Let $r_i^{(1)} = e^*R_ie$, $s_i = r_i^{(1)}I_k$, where I_k denotes the kth-order identity. Then, for all i,*

$$R_i^*B_{i+1} = B_is_i, \tag{5.2}$$

$$B_{i+1}^*R_i = s_iB_i^*. \tag{5.3}$$

Proof. Since the diagonal elements of R_i are positive, $[r_i^{(1)}]^* = r_i^{(1)}$, $s_i^* = s_i$. Since (5.3) follows immediately from (5.2), we need prove only (5.2). Now,

$$R_i^*B_{i+1} = (R_i^*e, R_i^*A_{i+1}e, \ldots, R_i^*A_{i+1}^{k-1}e).$$

From the definition of QR transform we can easily show that

$$R_i^*A_{i+1} = A_iR_i^*.$$

Hence

$$R_i^*B_{i+1} = (R_i^*e, A_iR_i^*e, \ldots, A_i^{k-1}R_i^*e).$$

An application of Lemma 3.2 with $k = 1$ yields

$$R_i^*B_{i+1} = (er_i^{(1)}, A_ier_i^{(1)}, \ldots, A_i^{k-1}er_i^{(1)}).$$

Hence

$$R_i^*B_{i+1} = B_ir_i^{(1)}I_k,$$

which completes the proof.

Let

$$u_i(\lambda) = B_i^*V_i(\lambda)B_i. \tag{5.4}$$

The transformation rule for $u_i(\lambda)$ is given by

THEOREM 5.1.

$$u_{i+1}(\lambda) = \frac{u_i(\lambda, \sigma_i, \sigma_i)}{[u_i^{(1)}(\sigma_i)]'}, \qquad u_i^{(1)}(\sigma_i) = e^*(A_i - \sigma_i)^{-1}e.$$

Proof. From (3.4),

$$V_{i+1}(\lambda) = R_i V_i(\lambda, \sigma_i, \sigma_i) R_i^*.$$

Hence, by Lemma 5.1,

$$u_{i+1}(\lambda) = B_{i+1}^* R_i V_i(\lambda, \sigma_i, \sigma_i) R_i^* B_{i+1} = s_i^2 B_i^* V_i(\lambda, \sigma_i, \sigma_i) B_i$$
$$= |r_i^{(1)}|^2 u_i(\lambda, \sigma_i, \sigma_i).$$

Taking $k = 1$ in (3.12) yields the result.

Let

$$b_i = B_i^* A_i(\lambda) B_i. \tag{5.5}$$

The transformation rule for b_i is given by

THEOREM 5.2.

$$b_{i+1} = \frac{\sigma_i u_i'(\sigma_i) + u_i(\sigma_i)}{[u_i^{(1)}(\sigma_i)]'}.$$

Proof. We have

$$A_{i+1} - \sigma_i = R_i V_i(\sigma_i) R_i^*.$$

Hence

$$b_{i+1} - \sigma_i B_{i+1}^* B_{i+1} = B_{i+1}^* R_i V_i(\sigma_i) R_i^* B_{i+1} = s_i^2 B_i^* V_i(\sigma_i) B_i.$$

Thus

$$b_{i+1} - \sigma_i B_{i+1}^* B_{i+1} = |r_i^{(1)}|^2 u_i(\sigma_i). \tag{5.6}$$

Now

$$B_{i+1}^* B_{i+1} = B_{i+1}^* R_i R_i^{-1} Q_i^{-1} Q_i R_i^{-*} R_i^* B_{i+1} = s_i^2 B_i^* V_i^2(\sigma_i) B_i.$$

Hence

$$B_{i+1}^* B_{i+1} = |r_i^{(1)}|^2 u_i'(\sigma_i) \tag{5.7}$$

and

$$b_{i+1} = [r_i^{(1)}]^2[\sigma_i u_i'(\sigma_i) + u_i(\sigma_i)]. \tag{5.8}$$

Therefore, by (3.12) with $k = 1$, the theorem follows.

THEOREM 5.3. *All poles of $u_i(\lambda)$ are eigenvalues of A. Furthermore $u_i(\lambda)$ has at least one pole.*

Proof. The proof is analogous to the proof of Theorem 4.1.

THEOREM 5.4. *A Class Two shift σ_{i+1} satisfies*

$$\det[\sigma_i - \sigma_{i+1} + u_i(\sigma_i)[u_i'(\sigma_i)]^{-1}] = 0. \tag{5.9}$$

Proof. The defining equation for a Class Two shift may be written as

$$\det[b_{i+1} - \lambda B_{i+1}^* B_{i+1}] = 0.$$

From (5.7) and (5.8),

$$b_{i+1} - \lambda B_{i+1}^* B_{i+1} = [r_i^{(1)}]^2[(\sigma_i - \lambda)u_i'(\sigma_i) + u_i(\sigma_i)]$$

and the result follows.

From Theorems 5.1 and 5.4 we can summarize as follows. The QR algorithm with Class Two shifts is the iteration

$$\det[\sigma_i - \sigma_{i+1} + u_i(\sigma_i)[u_i'(\sigma_i)]^{-1}] = 0, \qquad u_{i+1}(\lambda) = \frac{u_i(\lambda, \sigma_i, \sigma_i)}{[u_i^{(1)}(\sigma_i)]'}. \tag{5.10}$$

6. CLASS TWO SHIFTS ARE ZEROS OF LANCZOS POLYNOMIALS

Let

$$\alpha_i^{(j)} = e^* A_i^j e, \qquad j = 0, 1, \ldots, \tag{6.1}$$

be the Schwarz constants of A_i. Let $h_i^{(k)} = h_i$ denote the Hankel matrix

$$h_i = \begin{bmatrix} \alpha_i^{(0)}, & \ldots & \alpha_i^{(k-1)} \\ & \vdots & \\ \alpha_i^{(k-1)}, & \ldots & \alpha_i^{(2k-2)} \end{bmatrix}. \tag{6.2}$$

Let $H_i^{(k)} = H_i = \det h_i$ and

$$H_i^{(k)}(\lambda) = H_i(\lambda) = \det \begin{bmatrix} \alpha_i^{(0)}, & \ldots, & \alpha_i^{(k)} \\ & \vdots & \\ \alpha_i^{(k-1)}, & \ldots, & \alpha_i^{(2k-1)} \\ 1, & \ldots, & \lambda^k \end{bmatrix}. \tag{6.3}$$

The Lanczos polynomial $\chi_i^{(k)}(\lambda) = \chi_i(\lambda)$ is defined [3, p. 23] as the monic polynomial

$$\chi_i^{(k)}(\lambda) = \chi_i(\lambda) = \frac{H_i(\lambda)}{H_i}. \tag{6.4}$$

We have

THEOREM 6.1. *Class Two shifts are zeros of Lanczos Polynomials.*

Proof. Let $B_i = (e, A_i e, \ldots, A_i^{k-1}e)$. Then

$$\begin{aligned}
&\det[B_i^*(A_i - \lambda)B_i] \\
&= \det \begin{bmatrix} \alpha_i^{(1)} - \alpha_i^{(0)}\lambda, & \ldots, & \alpha_i^{(k)} - \alpha_i^{(k-1)}\lambda \\ & \vdots & \\ \alpha_i^{(k)} - \alpha_i^{(k-1)}\lambda, & \ldots, & \alpha_i^{(2k-1)} - \alpha_i^{(2k-2)}\lambda \end{bmatrix} \\
&= (-1)^k \det \begin{bmatrix} \alpha_i^{(0)}, & \alpha_i^{(1)} - \alpha_i^{(0)}\lambda, & \ldots, & \alpha_i^{(k)} - \alpha_i^{(k-1)}\lambda \\ & & \vdots & \\ \alpha_i^{(k-1)}, & \alpha_i^{(k)} - \alpha_i^{(k-1)}\lambda, & \ldots, & \alpha_i^{(2k-1)} - \alpha_i^{(2k-2)}\lambda \\ 1, & 0, & \ldots, & 0 \end{bmatrix} \\
&= (-1)^k \det \begin{bmatrix} \alpha_i^{(0)}, & \alpha_i^{(1)}, & \ldots, & \alpha_i^{(k)} \\ & & \vdots & \\ \alpha_i^{(k-1)}, & \alpha_i^{(k)}, & \ldots, & \alpha_i^{(2k-1)} \\ 1, & \lambda, & \ldots, & \lambda^k \end{bmatrix} \\
&= (-1)^k H_i \chi_i(\lambda).
\end{aligned}$$

Since $H_i = \det(B_i^* B_i)$, which is nonzero by assumption, the theorem is proved.

7. GLOBAL CONVERGENCE

We shall prove the following theorem.

THEOREM 7.1. *Let A be Hermitian. Then QR with second-order Class Two shifts is convergent.*

Before proving this theorem we introduce some new notation and then prove a number of inequalities.

Rather than dealing with the B_i matrix which has linearly independent columns, it is convenient to introduce a matrix with orthogonal columns which span the same subspace. We can write

$$B_i = \tilde{B}_i U_i,$$

where U_i is a unit upper-triangular matrix and $\tilde{B}_i^* \tilde{B}_i$ is diagonal. Then

$$\det[B_i^*(A_i - \lambda)B_i] = \det[\tilde{B}_i^*(A_i - \lambda)\tilde{B}_i].$$

Let

$$\tilde{B}_i = (b_i^{(0)}, b_i^{(1)}, \ldots, b_i^{(k-1)}).$$

Then

$$b_i^{(j)} = \chi_i^{(j)}(A_i)e, \qquad j = 0, 1, \ldots, k-1, \tag{7.1}$$

where $\chi_i^{(j)}(\lambda)$ is the Lanczos polynomial defined by (6.4). A proof of (7.1) in a somewhat different setting may be found in Householder [3, p. 22]. Define

$$b_i^{(k)} = \chi_i^{(k)}(A_i)e.$$

The symbol $\|\cdot\|$ denotes the Euclidean vector norm. We have, for all monic polynomials $\varphi^{(j)}$ of degree j, $j = 0, 1, \ldots, k$,

$$\|b_i^{(j)}\| = \min_{\varphi^{(j)}} \|\varphi^{(j)}(A_i)e\|, \tag{7.2}$$

equality being attained for $\varphi^{(j)} = \chi_i^{(j)}$. See Lanczos [5].

From (6.4),

$$\chi_i^{(1)}(\lambda) = \lambda - \alpha_i^{(1)}. \tag{7.3}$$

By Theorem 6.1, the shift satisfies $\chi_i^{(2)}(\sigma_i) = 0$. Let

$$\chi_i^{(2)}(\lambda) = (\lambda - \sigma_i)(\lambda - \omega_i). \tag{7.4}$$

Define

$$\beta_i = \|b_i^{(1)}\|, \qquad \gamma_i = \|b_i^{(2)}\|. \tag{7.5}$$

As earlier,

$$V_i(\lambda) = (A_i - \lambda)^{-1}, \qquad r_i^{(1)} = r_i = e^* R_i e.$$

From (3.12), with $k = 1$,

$$\|V_i(\sigma_i)e\| = r_i^{-2}. \tag{7.6}$$

We now prove a number of inequalities which are required in the proof of the theorem.

LEMMA 7.1.

$$r_i^2 \leqslant \frac{2\beta_i^2\gamma_i^2}{\gamma_i^2 + 2\beta_i^4}. \tag{7.7}$$

Proof. We write the vector $V_i(\sigma_i)e$ as a linear combination of $b_i^{(0)}$, $b_i^{(1)}$, $b_i^{(2)}$, and an orthogonal component which we label d_i. Thus

$$V_i(\sigma_i)e = \sum_{k=0}^{2} g_i^{(k)} b_i^{(k)} + d_i,$$

$$g_i^{(k)} = \frac{[b_i^{(k)}]^* V_i(\sigma_i)e}{[b_i^{(k)}]^* b_i^{(k)}}, \qquad k = 0, 1, 2. \tag{7.8}$$

Let

$$v_i(\lambda) = e^* V_i(\lambda)e.$$

Then

$$g_i^{(0)} = v_i(\sigma_i),$$

$$g_i^{(1)} = \frac{e^*(A_i - \alpha_i^{(1)})V_i(\sigma_i)e}{\beta_i^2} = \frac{1 + (\sigma_i - \alpha_i^{(1)})v_i(\sigma_i)}{\beta_i^2},$$

$$g_i^{(2)} = \frac{e^*(A_i - \omega_i)e}{\gamma_i^2} = \frac{\alpha_i^{(1)} - \omega_i}{\gamma_i^2}.$$

Neglecting the contribution of $\|d_i\|$ and using (7.6), we find

$$\|V_i(\sigma_i)e\| = r_i^{-2} \geqslant v_i^2(\sigma_i) + \frac{[1 + (\sigma_i - \alpha_i^{(1)})v_i(\sigma_i)]^2}{\beta_i^2} + \frac{(\alpha_i^{(1)} - \omega_i)^2}{\gamma_i^2}$$

$$= \frac{1}{\beta_i^2}\{[\beta_i^2 + (\sigma_i - \alpha_i^{(1)})^2]v_i^2(\sigma_i) + 2(\sigma_i - \alpha_i^{(1)})v_i(\sigma_i) + 1\}$$

$$+ \frac{1}{\gamma_i^2}(\alpha_i^{(1)} - \omega_i)^2.$$

Finding the minimum, as a function of $v_i(\sigma_i)$, of the expression in braces, we have

$$r_i^{-2} \geqslant \frac{1}{\beta_i^2 + (\sigma_i - \alpha_i^{(1)})^2} + \frac{(\alpha_i^{(1)} - \omega_i)^2}{\gamma_i^2}. \tag{7.9}$$

Using (6.3) and (6.4),

$$\chi_i^{(2)}(\alpha_i^{(1)}) = [\alpha_i^{(1)}]^2 - \alpha_i^{(2)}.$$

Also,

$$\beta_i^2 = \|b_i^{(1)}\|^2 = e^*(A_i - \alpha_i^{(1)})^2 e = \alpha_i^{(2)} - (\alpha_i^{(1)})^2.$$

Hence

$$\chi_i^{(2)}(\alpha_i^{(1)}) = (\alpha_i^{(1)} - \sigma_i)(\alpha_i^{(1)} - \omega_i) = -\beta_i^2.$$

Since σ_i is the zero of $\chi_i^{(2)}$ closest to $\alpha_i^{(1)}$, we conclude that

$$|\alpha_i^{(1)} - \sigma_i| \leqslant \beta_i, \qquad |\alpha_i^{(1)} - \omega_i| \geqslant \beta_i. \tag{7.10}$$

Using (7.10) in (7.9),

$$r_i^{-2} \geqslant \frac{1}{2\beta_i^2} + \frac{\beta_i^2}{\gamma_i^2}$$

and the result follows.

LEMMA 7.2.

$$\beta_{i+1}^2 \leqslant r_i^2 \leqslant \beta_i^2 \frac{2\gamma_i^2}{\gamma_i^2 + 2\beta_i^4}. \tag{7.11}$$

Proof. The inequality on the right was obtained in the Lemma 7.1. By definition,

$$\beta_{i+1} = \|(A_{i+1} - \alpha_{i+1}^{(1)})e\|.$$

By the minimum property (7.2),

$$\begin{aligned}\|(A_{i+1} - \alpha_{i+1}^{(1)})e\| &\leqslant \|(A_{i+1} - \sigma_i)e\| = \|Q_i{}^*R_i{}^*e\| \\ &= \|Q_i{}^*er_i{}^*\| = \|Q_i{}^*e\|\cdot\|r_i{}^*\| = r_i,\end{aligned}$$

which proves the left inequality.

LEMMA 7.3.

$$\gamma_{i+1}^2 \leqslant r_i{}^2\beta_i{}^2 \leqslant \gamma_i{}^2\,\frac{2\beta_i{}^4}{\gamma_i{}^2 + 2\beta_i{}^4}. \tag{7.12}$$

Proof. The inequality on the right follows from Lemma 7.1. By (7.1), (7.4), and (7.5),

$$\gamma_{i+1} = \|(A_{i+1} - \omega_{i+1})(A_{i+1} - \sigma_{i+1})e\|$$

and, by (7.2),

$$\begin{aligned}\gamma_{i+1} &\leqslant \|(A_{i+1} - \alpha_i{}^{(1)})(A_{i+1} - \sigma_i)e\| = \|(Q_i{}^*R_i{}^* + \sigma_i - \alpha_i{}^{(1)})(Q_i{}^*R_i{}^*e)\| \\ &= \|Q_i{}^*(A_i - \alpha_i{}^{(1)})R_i{}^*e\| = \|(A_i - \alpha_i{}^{(1)})e\|\cdot\|r_i{}^*\| = \beta_i r_i.\end{aligned}$$

Hence

$$\gamma_{i+1} \leqslant \beta_i r_i,$$

and the result follows.

LEMMA 7.4.

$$\beta_{i+1}\gamma_{i+1} \leqslant \frac{\beta_i\gamma_i}{\sqrt{2}}, \tag{7.13}$$

$$\beta_{i+1}^3 \leqslant \beta_i\gamma_i. \tag{7.14}$$

Proof. From Lemmas 7.2 and 7.3,

$$\beta_{i+1}\gamma_{i+1} \leqslant \beta_i\gamma_i\,\frac{2\beta_i{}^2\gamma_i}{\gamma_i^2 + 2\beta_i{}^4}.$$

Since

$$\gamma_i^2 + 2\beta_i^4 \geqslant \gamma_i^2 + 2\beta_i^4 - (\gamma_i - \sqrt{2}\beta_i^2)^2 = 2\sqrt{2}\beta_i^2\gamma_i, \tag{7.15}$$

we conclude that

$$\beta_{i+1}\gamma_{i+1} \leqslant \frac{\beta_i\gamma_i}{\sqrt{2}},$$

which proves (7.13).

From Lemma 7.2, $\beta_{i+1} \leqslant \sqrt{2}\beta_i$. Then another application of Lemma 7.2 yields

$$\beta_{i+1}^3 \leqslant 2\sqrt{2}\beta_i\gamma_i \frac{\beta_i^2\gamma_i}{\gamma_i^2 + 2\beta_i^4} \leqslant \beta_i\gamma_i,$$

which completes the proof.

We are now ready to give a

Proof of Theorem 7.1. By (7.13), the sequence of $\beta_i\gamma_i$ converges to zero. By (7.14), the sequence of β_i converges to zero. That is, $\|(A_i - \alpha_i^{(1)})e\|$ converges to zero, and this implies convergence.

REFERENCES

1 F. L. Bauer, Numerische Abschätzung und Berechnung von Eigenwerten nicht-symmetrischer Matrizen, *Aplikace Mat.* **10**(1965), 178–189.

2 J. G. F. Francis, The QR transformation, *Computer J.*, **4**(1961), 265–271, 332–345.

3 A. S. Householder, *The Theory of Matrices in Numerical Analysis*, Blaisdell, Waltham, Mass., 1964.

4 M. A. Jenkins and J. F. Traub, A three-stage variable-shift iteration for polynomial zeros and its relation to generalized Rayleigh iteration, *Num. Math.* **14**(1970), 252–263.

5 C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Standards*, **45**(1950), 255–282.

6 B. N. Parlett, Global convergence of the basic QR algorithm on Hessenberg matrices, *Math. Comp.* **22**(1968), 803–817.

7 B. N. Parlett and W. Kahan, On the convergence of a practical QR algorithm, *Proc. IFIP Congr.*, 1968, pp. 25–30.

8 J. H. Wilkinson, *The Algebraic Eigenvalue Problem*, Oxford University Press, New York, 1965.

9 J. H. Wilkinson, Global convergence of tridiagonal QR algorithm with origin shifts, *Linear Algebra*, **1**(1968), 409–420.

Received January, 1969